The Patent Office

INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed Evans.

Dated 24 February 2000

THE PATENT OFFICE
F
27 NOV 1999
NEWPORT

The Patent Office

7/77

Patents Act 1977
Rule 15

Statement of inventorship and of right to grant of a patent

The Patent Office

Concept House
Cardiff Road
Newport
South Wales NP9 1RH

27 NOV 1999

1.	Your reference	UK999107
2.	Patent application number *(if you know it)*	9928011.7
3.	Full name of the or of each applicant	INTERNATIONAL BUSINESS MACHINES CORPORATION
4.	Title of invention	Voice Processing System
5.	State how the applicant(s) derived the right from the inventor(s) to be granted a patent	By employment and by agreement
6.	How many, if any, additional Patents Forms 7/77 are attached to this form?	
7.		I/We believe that the person(s) named over the page (and on any extra copies of this form) is/are the inventor(s) of the invention which the above patent application relates to.

Signature

26 November 1999
Date

8.	Name and daytime telephone number of person to contact in the United Kingdom	S R Davies 01962 816369

Enter the full names, addresses and postcodes of the inventors in the boxes and underline the surnames

John Brian PICKERING
(UK resident)
c/o IBM United Kingdom Limited
Intellectual Property Law
Hursley Park
Winchester
Hampshire
SO21 2JN
UK

Patents ADP number *(if known)*

7119970001

Patents ADP number *(if known)*

If there are more than three inventors, please write their names and addresses on the back of another Patents Form 7/77 and attach it to this form

Patents ADP number *(if known)*

REMINDER

Have you signed the form?

Patents Form 1/77

THE PATENT OFFICE
F
27 NOV 1999
NEWPORT

Patents Act 1977
Rule 16

The Patent Office

1/77

29NOV99 E495067-1 C57465
F01/7700 0.00-9928011.7

Request for grant of a patent

The Patent Office

Concept House
Cardiff Road
Newport
South Wales NP9 1RH

27 NOV 1999

1.	Your reference	UK999107
2.	Patent application number *(The Patent Office will fill in this part)*	**9928011.7**
3.	Full name, address and postcode of the or of each applicant *(underline all surnames)*	INTERNATIONAL BUSINESS MACHINES CORPORATION Armonk New York 10504 United States of America
	Patents ADP number *(if you know it)*	51963702
	If the applicant is a corporate body, give the country/state of its incorporation	State of New York United States of America
4.	Title of the invention	Voice Processing System
5.	Name of your agent *(if you have one)*	S R Davies
	"Address for Service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*	IBM United Kingdom Limited Intellectual Property Department Hursley Park Winchester Hampshire SO21 2JN
	Patents ADP number *(if you know it)*	

6.	If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number	Country	Priority App No *(if you know it)*	Date of filing *(day/month/year)*
7.	If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application	No of earlier application		Date of filing *(day/month/year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*
a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.)

Yes

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	11
Claim(s)	2
Abstract	1
Drawing(s)	2+2

10. If you are also filing any of the following, state how many against each item.

Priority documents	-
Translations of priority documents	-
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	2
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application

Signature

26 November 1999
Date

12. Name and daytime telephone number of person to contact in the United Kingdom

S R Davies
01962 816369

VOICE PROCESSING SYSTEM

The present invention relates to voice processing systems and the like, and more particularly to voice processing systems that use speech recognition.

Voice processing systems whereby callers interact over the telephone network with computerised equipment are very well-known in the art, and include voice mail systems, voice response units, and so on. Typically such systems ask a caller (or called party) questions using prerecorded prompts, and the caller inputs answers by pressing dual tone multiple frequency (DTMF) keys on their telephones. This approach has proved effective for simple interactions, but is clearly restricted in scope due to the limited number of available keys on a telephone. For example, alphabetical input is particularly difficult using DTMF keys.

There has therefore been an increasing tendency in recent years for voice processing systems to use speech recognition in order to augment DTMF input (n.b. the terms speech recognition and voice recognition are used interchangeably herein to denote the act of converting a spoken audio signal into text). The utilisation of speech recognition permits the handling of callers who do not have a DTMF phone, and also the acquisition of more complex information beyond simple numerals from the caller.

As an illustration of the above, WO96/25733 describes a voice response system which includes a prompt unit, a Voice Activity Detector (VAD), and a voice recognition unit. In this system, as a prompt is played to the caller, any input from the caller is passed to the VAD, together with the output from the prompt unit. This allows the VAD to perform echo cancellation on the incoming signal. Then, in response to the detection of voice by the VAD, the prompt is discontinued, and the caller input is switched to the recognition unit, thereby providing a barge-in facility.

Speech recognition in a telephony environment can be supported by a variety of hardware architectures. Many voice processing systems include a special DSP card for running speech recognition software. This card is connected to a line interface unit for the transfer of telephony data by a time division multiplex (TDM) bus. Most commercial voice processing systems, more particularly their line interface units and DSP cards, conform to one of two standard architectures: either the Signal Computing System Architecture (SCSA), or the Multi-vendor Integration Protocol (MVIP). A somewhat different configuration is described in GB 2280820, in which a voice processing system is connected via a local area network to a remote server, which provides a speech recognition facility. This

approach is somewhat more complex than the TDM approach, given the data communication and management required, but does offer significantly increased flexibility.

Speech recognition systems are generally used in telephony environments as cost-effective substitutes for human agents, and are adequate for performing simple, routine tasks. It is important that such tasks are performed accurately, otherwise there may be significant caller dissatisfaction, and also as quickly as possible, both to improve caller throughput, and because the owner of the voice processing system is often paying for the call via some FreePhone mechanism (eg an 0800 number), or because an outbound application is involved.

(Note that as used herein, the term "caller" simply indicates the party at the opposite end of a telephone connection to the voice processing system, rather than to specify which party actually initiated the telephone connection).

One facility in prior art voice processing systems to help accelerate call handling and also to improve the user interface is barge-in. As briefly indicated above, this is where voice recognition is enabled on an incoming channel at the same time as the system is playing a prompt on the corresponding outgoing channel. This allows a caller to interrupt the prompt as soon as they know what response to give. For example, if the prompt is "Say Account for account information, say Order to order material, or say Transfer to speak to an operator", and the caller wants account information, barge-in allows the caller to interrupt the prompt by saying "Account" before the complete prompt has finished. This is particularly useful for regular callers who are familiar with the application and the prompt menus. Following such an interruption, the application abandons the rest of the prompt and the caller interruption is passed to the recognition system for processing. The application can then proceed further on the basis of what is returned from the recognition system.

One problem with prior art barge-in systems is that they can be confused by noise on the telephone line. For example, if the caller coughs, the outgoing prompt may be suspended even though the caller actually still desires to hear the rest of the prompt. This can leave a very awkward situation, with the machine expecting further input from the caller, and the caller expecting further output from the machine. The result can be a suspended or confused dialogue with the caller, resulting ultimately in a wasted or highly ineffective call.

A known solution to this for discrete word (small vocabulary) recognition systems, which typically only recognise one or two dozen

different inputs (eg numerals 0-9), is to wait for the recognition result to be returned before interrupting the outgoing prompt. Thus if the supposed caller input is not recognised, perhaps because it is noise or some irrelevant caller interjection, then the playout of the prompt is continued. In other words, the prompt is only interrupted where there is a successful recognition result.

Although this approach which essentially involves modelling the recognition system and application to the likely range of caller responses is effective for discrete word systems, more modern voice processing applications often involve large vocabulary speech recognition for which such modelling is not feasible. For these applications, the provision of barge-in is prone to trigger the termination the prompt even in circumstances where this was not actually the intention of the caller.

Accordingly, the invention provides a method of providing speech recognition with barge-in for a voice processing system comprising the steps of:

playing out a prompt to a user;

receiving audio input from the user whilst said prompt is still being played out;

performing speech recognition on said audio input to determine a corresponding text;

performing lexical analysis on said text to determine whether or not the text satisfies one or more conditions;

and responsive to said text satisfying said one or more conditions, terminating the playing out of the prompt; otherwise, continuing the playing out of said prompt.

The invention allows the voice processing system to assess whether or not to give effect to barge-in, and interrupt the outgoing prompt, based on an analysis of what has actually been said. In particular, this allows the system to discriminate some background comment, whether by the user or someone with them, from a positive response to the prompt. Thus where the input is determined to be not relevant to the prompt, the playout of the prompt is continued, and the user is not left accidentally suspended in an application script. In this case the recognised text can generally be discarded as irrelevant to the actual dialogue between the user and the voice processing system.

In the preferred embodiment the lexical analyser determines whether or not the response is relevant to the prompt by scanning the text to see if it contains one or more predetermined words (note that the list of possible words to be matched may vary with each prompt). For instance, if a US caller is asked to name the state in which they live, the lexical

analyser may simply be provided with a list of 50 states to match dynamically.

Although this positive matching of text with a predefined set of words is the probably the most convenient approach to take, in other cases a different strategy may be appropriate. For example, if a caller is being asked to input their name, and the lexical analyser determines that first two words they input are both dictionary words, it may consider it most probable that the response so far does not in fact represent their name, and so not give effect to barge-in.

In some embodiments, one or more acoustic parameters may also be used as a secondary measure for determining whether to continue or to terminate playing out of said prompt. For example, it may be desirable to refrain from effecting barge-in if the audio input is muted or highly variable in volume, this being taken as an indication that the input is not in fact intended as an answer to the prompt.

Typically the voice processing system and the user or caller communicate with each other over a telephone network. Thus the prompt is played out over a telephone connection, and the audio input is received back over the telephone connection.

The invention further provides a voice processing system for providing speech recognition with barge-in, said voice processing system comprising:

means for playing out a prompt to a user;

means for receiving audio input from the user whilst said prompt is still being played out;

means for performing speech recognition on said audio input to determine a corresponding text;

means for performing lexical analysis on said text to determine whether or not the text satisfies one or more conditions;

and means responsive to said text satisfying said one or more conditions, for terminating the playing out of the prompt; otherwise, for continuing the playing out of said prompt.

Such a voice processing system may be adapted for connection to the telephone network (conventional PSTN or the Internet), in a customer server kiosk, or in any other appropriate device. Note that the speech recognition means and/or the lexical analysis may or may not be integral to the voice processing system itself (as will be apparent more clearly from the preferred embodiments described below).

The invention further provides a computer readable medium containing computer program instructions for a voice processing system

for providing speech recognition with barge-in, said computer program instructions comprising instructions for:

playing out a prompt to a user;

receiving audio input from the user whilst said prompt is still being played out;

performing speech recognition on said audio input to determine a corresponding text;

performing lexical analysis on said text to determine whether or not the text satisfies one or more conditions;

and responsive to said text satisfying said one or more conditions, terminating the playing out of the prompt; otherwise, continuing the playing out of said prompt.

The computer readable medium may comprise a magnetic or optical disk, solid state memory device, tape, or other appropriate storage apparatus. In some cases this medium may be physically loadable into the storage device. In other cases, this medium may be fixed in the voice processing system, and the instructions loaded onto the medium via some wired or wireless network connection. Another possibility is for the medium to be remote from the voice processing system itself, with the instructions being downloaded over a wired or wireless network connection for execution by the voice processing system.

It will be appreciated that the computer program and apparatus of the invention will benefit from substantially the same preferred features as the method of the invention.

An embodiment of the invention will now be described in detail by way of example only with reference to the following drawings:

Figure 1 is a simplified schematic diagram of a voice processing system connected to a remote server; and

Figure 2 is a schematic flow chart illustrating the method of operation of the system of Figure 1 in a preferred embodiment of the invention.

Figure 1 illustrates in simplified schematic form the main hardware and software components of a voice processing system 10 having speech recognition functionality. The system 10 is connected to the telephone network 110 by one or more telephony channels 150. In Figure 1 these are digital trunk lines, each carrying multiple telephony channels (T1 digital trunk lines in North America comprise 24 channels, E1 digital trunk lines in Europe comprise 30 channels). It will be appreciated that in some installations, the digital trunk lines may be connected to the telephone network 110 via a switch or PBX; indeed one possibility is for the voice processing system to be a part of the telephone network itself, providing intelligent services to the network.

The system unit 145 of the voice processing system comprises a computer workstation, which runs an appropriate operating system 200 and a voice processing program 210, which in turn interacts with a user application 220 in order to determine how calls should be processed. The voice processing platform also supports a lexical analyser 245, the purpose of which will be described in more detail below. The system unit includes one or more digital trunk adapter cards 160 for interfacing to the telephone network 110 via link 150. The system unit also contains a network interface card 170 which allows programs running on the workstation to communicate over a local area network (LAN) 250, using communications software 205 incorporated into the operating system. In addition, the system unit includes a digital signal processing (DSP) card 180, which is connected to the trunk adapter (or adapters) via a time division multiplex (TDM) bus 165. Speech recognition software 185 may be installed as microcode or firmware on the DSP card.

The adapter cards 160 are responsible for interfacing with the telephone network over their respective lines, including signalling, demultiplexing incoming telephony signals, and so on. They may also be used to perform activities such as voice activity detection on a particular telephony channel, compression/decompression of voice signals, and DTMF recognition and generation, although these may be performed on the DSP card instead.

A server system 300 is attached to the LAN 250 via network interface card 310, and supports an operating system and appropriate communications software (not shown), and speech recognition software 320. It will be appreciated therefore that there are two voice recognition resources available to the application 220, the first of these being locally installed on the DSP card 180 in the system unit, and the second voice recognition resource being available remotely via the LAN 250 on server 300.

In one preferred embodiment, the voice processing system is the IBM Voice Response unit (previously known as the DirectTalk voice processing system) available from the IBM Corporation, running on an RS/6000 workstation on top of the AIX operating system. The voice recognition resource comprises a large vocabulary voice recognition system and may, for example, be the ViaVoice engine, available from IBM Corporation. PC-based systems are also available.

It will be appreciated that there are many possible variations in the design of the voice processing system of Figure 1. For example, some voice processing systems accept input from analog lines rather than digital trunks, whilst other voice processing systems are connected to the Internet instead of, or as well as, the conventional telephone

network (this provides Voice over IP capability). Some voice processing systems may package the DSP card 185 as a daughter card of the digital trunk adapter 160, thereby eliminating the need for the TDM bus. In addition, the illustrated system has access both to a server voice recognition system, and also to a local DSP recognition system, whereas many voice processing systems will have access to only one such resource. Further, any suitable network could be used for communications between the server and the voice processing system, providing it has suitable transmission characteristics in terms of bandwidth and latency (eg one possibility might be to use an ATM connection). Additionally, although the voice processing system illustrated has just a single local DSP resource, some voice processing systems may include multiple DSP cards, with each card supporting multiple recognition programs running simultaneously. Moreover, although the server approach as shown has the recognition system installed on a separate machine from the line interface unit 160, it would clearly be possible for the software recognition system to be running on the same machine 145 as the line interface unit, provided this machine had sufficient processing capability. Conversely, the lexical analyser could run on a separate system from system unit 145, for example, on system 300, or possibly on some other system (not shown) connected to LAN 250. The skilled person will appreciate that such variations are not pertinent to the principles of the present invention.

A typical operation of the voice processing system of Figure 1 is as follows. An incoming call is received over the telephony network 110, and the voice processing system 10, under the control of the user application 220, may determine that the caller does not have a DTMF telephone (this determination can be made for example through knowledge of the calling or called number, or by the caller failing to depress a tone key when requested to do so). Alternatively, the machine may be configured to use voice recognition for all callers. In any event, the voice processing system proceeds to utilise voice recognition to interpret the caller's input. This is done by forwarding the appropriate audio signal to a speech recognition resource. For use of the local resource, this leads to a channel on the trunk adapter 160 being connected with a channel on the DSP card 180 via the TDM bus 165. Access to the remote resource can be achieved for example using the approach described in GB 2325110, which is incorporated herein by reference. (Note that the TDM bus connection as shown in Figure 1 provides a uni-directional data flow, so that as is well-known in the art, a pair of such connections are needed to support a full duplex telephone conversation).

The speech recognition system processes the received audio signal, and returns the spoken string to the user application 220. It will be

noted that the large majority of voice recognition systems used in telephony are speaker independent; in other words, such systems are trained to recognise vocal input from any speaker. This is accomplished by training the machine on vocal samples from a representative set of speakers. In general the recognition may also return other information in addition to the recognised string, for example, a statistical confidence level, possible alternative results, and so on. The user application will then further process the call in accordance with the returned recognition result. For example, if the caller has spoken a numerical sequence corresponding to an account number, it may provide the caller with the ability to access account information; if the caller has spoken a particular name, it may transfer the caller to the telephone extension for that name. If the recognition fails, the caller may be asked to repeat the input, or transferred to a human operator.

The voice processing system of Figure 1 supports barge-in, in that the recognition system 185, 320 can be active for a line at the same time that a prompt is being played out on that line. This operation is shown in Figure 2, where the application first enables barge-in (step 400), and then starts to play out a prompt to the caller (step 410). (It will be appreciated that in some circumstances it may not be desirable to permit caller barge-in; thus the voice processing software 210 typically permits the application 220 to enable/disable this on a per prompt basis).

The method then proceeds to step 420, in which a test is made to see if the caller has said anything. This test is generally performed by a voice activity detector, so that background tones and other extraneous noise will be discounted. If no spoken caller input is detected, then the playout of the prompt is continued (step 450), until the end of the prompt is reached (step 455). This effectively represents the end of the period of potential barge-in. The system then waits for caller input in conventional fashion (step 470). What happens next is not of direct interest to the present invention (hence the exit (step 480) in Figure 2), but generally the caller will make some input which will be passed to the recognition unit; the voice processing system will then typically take some further action based on the recognised caller input. If no caller input is received within a predetermined period, the system will time out (not shown). The application may then choose perhaps to transfer the caller to a human agent, or to replay them the prompt; similar possibilities are also typically provided if the recognition on the caller input fails.

Returning now to step 430, this is reached where some spoken input from the caller is detected at step 420. The received voice signal is passed to the recognition device; note that at this stage the prompt play-out continues. Once a response has been received from the

recognition device, this is then passed to the lexical analyser (step 435); note again that the prompt play-out is still continuing at this point. (If the recognition fails, so that no recognition text is produced, the method can go direct from step 430 to step 450, effectively as per the prior art).

The lexical analyser tests the recognised text against a predetermined set of words or criteria (step 440) which will be described in more detail below. If the result of this test is positive, then the method proceeds to step 460, where output of the prompt is terminated. The remainder of the caller input is then detected (step 465), whereupon processing can proceed as described above, with the method terminating at step 480.

Thus a form of barge-in has been effected. This is distinguished from that of the prior art, in that such barge-in occurs only when the lexical analyser confirms that the recognised text is relevant to the needs of the application. In other words, barge-in requires not only that the caller interjection be successfully recognised, but also that it passes whatever criterion (or criteria) the lexical analyser applies. This second step is particularly important, since the increasing use of large vocabulary systems in voice processing applications makes it ever more likely that the system will recognise some background or other remark at the caller end which is not a direct answer to the prompt. By being able to discriminate against such remarks or utterances, the robustness of the dialogue between the caller and the voice processing system is greatly enhanced.

In the preferred embodiment, the lexical analyser is provided with a predefined set of task words. The lexical analyser then tests whether or not the recognised text includes one or more of the task words. If not, the caller input may not be relevant (yet) to the information required, and so the prompt play-out is allowed to continue. Thus the prompt play-out is only terminated when it is specifically confirmed that the caller input is likely to be a valid response to the prompt. Note that the predefined set of task words may vary from one prompt to another.

As an example, the first part of an application prompt may ask for an account code. A second part of the prompt may go on to state that the account code is printed top left on all statements. It is desirable to allow barge-in after the first part of the prompt, for those who already know their account code. In this case, the lexical analyser is set up to match only digits. This will then exclude barge-in being triggered in circumstances where (for example) the telephone picks up the caller shouting "Dear, what is our account code", or mumbling to themselves

"Where did I put this now". This avoids the prior art problem where barge-in is unintentionally activated by the caller, resulting in the dialogue being accidentally suspended in mid-air, with neither the caller nor the voice processing system knowing how to proceed.

Based on the above, the skilled person will be able to develop further possible criteria for use in the lexical analyser. One variation is that rather than look for single task words, the lexical analyser may look for particular word phrases or combinations, grammatical structures, etc. For example, in a product ordering application, a caller interruption starting: "I'd like" might trigger barge-in. Further, the lexical analyser test may be inclusive or exclusive (i.e. the result may be arranged to be positive or negative, given the presence of certain features in the caller response).

It will be appreciated that the skilled person will be aware of many possible variations on the embodiment described above. For example, some systems may allow the caller to interrupt a prompt by pressing a DTMF key. This would then be detected at step 420 in Figure 2, and the system could take whatever action is required. (Note that it is quite possible for the voice processing system to support input by both DTMF key and by speech, whichever the caller happens to choose). Another possibility is for step 420 to be eliminated altogether, and replaced by step 430 (the flow would then be from 410 to 430 to 435 to 440 to either 450 or 460 as appropriate, with a negative result from step 455 returning to step 430). In other words, all input would be passed to the recognizer; during silence, the recognized text would be blank, and this would not be considered a valid response, so that processing would then pass to step 450). This approach avoids the use of a voice activity detector, but ties up a recognition channel from the moment the outgoing prompt is started, rather than from the moment that a positive caller response is detected. Since recognition resources tend to be expensive, this approach is therefore likely to be more costly than the approach depicted in Figure 2.

One enhancement to the approach of Figure 2 is to use acoustic parameters in addition to the lexical analysis to determine the relevance of the caller input (and hence whether or not to effect barge-in). The analysis of the acoustic parameters can be performed in parallel with the speech recognition and/or the lexical analysis, typically by some DSP unit provided in the voice processing system (e.g. card 180 in Figure 1). One example of an acoustic parameter which can be considered is volume: if this is low or highly variable, then this may be an indication that the caller is speaking away from the microphone, and therefore not directly answering the prompt (hence barge-in should not be effected). There may also be some variation in timbre or other acoustic property

such as spectral distribution if the caller's mouth moves further away from the speaker, providing another indication that the caller is not directly answering the prompt.

A further possibility is to try to perform some form of speaker identification, for example by estimating the vocal tract of the speaker from the fundamental frequency and formants of the incoming audio signal, and checking that the audio input is indeed coming from the caller. Thus a change in speaker may indicate that the input is being picked up from a background or secondary speaker, and so is not intended as an answer to the prompt (i.e. no barge-in). There may also be multiple voices, the so-called "cocktail party" effect, whereby the system can try to eliminate background contributions, and focus solely on the input from the caller. Various techniques for speaker identification/recognition are discussed in "Fundamentals of Speech Signal Processing" by Saito and Nakata, Academic Press, 1985, and "Vowel Perception and Production" by Rosner and Pickering, Oxford University Press, 1994.

Note that such acoustic parameters in most cases will not be solely determinative of whether or not to provide barge-in, but can provide useful auxiliary information. For example, it may be decided to continue the prompt (i.e. to refrain from effecting barge-in) if there are no particular recognised task words, and if this is backed up by some acoustic indication of non-pertinent input, such as a change in volume. The use of acoustic parameters to help determine whether or not to effect barge-in will generally increase the cost and complexity of the voice processing system, but may enhance the robustness of the overall solution.

Reviewing now the system shown in Figure 1, it will be appreciated that the recognition system and/or the lexical analyser may not be included in the voice processing system of the invention, but rather may be supplied by some external componentry. Alternatively, the recognition system and/or the lexical analyser may be part of or integral to the voice processing system. The lexical analyser may be included as part of the user application 220. It is also possible for the barge-in facility (with or without the lexical analyser) to be essentially incorporated within a recognition system.

Furthermore, although the system of the present invention has been described primarily in terms of a telephony environment, it is not so limited. For example, it might be included in an automatic kiosk, or in an in-car navigation and control unit. It will therefore be recognised that the system of the present invention may find utility in a great variety of voice processing applications.

CLAIMS

1. A method of providing speech recognition with barge-in for a voice processing system comprising the steps of:

playing out a prompt to a user;

receiving audio input from the user whilst said prompt is still being played out;

performing speech recognition on said audio input to determine a corresponding text;

performing lexical analysis on said text to determine whether or not the text satisfies one or more conditions;

and responsive to said text satisfying said one or more conditions, terminating the playing out of the prompt; otherwise, continuing the playing out of said prompt.

2. The method of claim 1, further comprising the step, responsive to said step of continuing the playing out of said prompt, of discarding said text.

3. The method of claim 1 or 2, wherein said step of performing lexical analysis to determine whether or not the text satisfies one or more conditions comprises scanning the text to see if it contains one or more predetermined words.

4. The method of claim 3, wherein said one or more predetermined words are specific to the particular prompt being played out.

5. The method of any preceding claim, in which said voice processing system and said user communicate with each other over a telephone network, whereby the prompt is played out over a telephone connection, and said audio input is received back over the telephone connection.

6. The method of any preceding claim, further comprising the step of using one or more acoustic parameters of the audio input to assist determining whether to continue or to terminate playing out of said prompt.

7. A voice processing system for providing speech recognition with barge-in, said voice processing system comprising:

means for playing out a prompt to a user;

means for receiving audio input from the user whilst said prompt is still being played out;

means for performing speech recognition on said audio input to determine a corresponding text;

means for performing lexical analysis on said text to determine whether or not the text satisfies one or more conditions;

and means responsive to said text satisfying said one or more conditions, for terminating the playing out of the prompt; otherwise, for continuing the playing out of said prompt.

8. The voice processing system claim 7, further comprising the means, responsive step of continuing the playing out of said prompt, for discarding said text.

9. The voice processing system of claim 7 or 8, wherein said means for performing lexical analysis to determine whether or not the text satisfies one or more conditions comprises means for scanning the text to see if it contains one or more predetermined words.

10. The voice processing system of claim 9, wherein said one or more predetermined words are specific to the particular prompt being played out.

11. The voice processing system of any of claims 7 to 10, in which said voice processing system and said user communicate with each other over a telephone network, whereby the prompt is played out over a telephone connection, and said audio input is received back over the telephone connection.

12. The voice processing system of any of claims 7 to 11, wherein said means for receiving caller input includes a voice activity detector for discriminating between speech input and other forms of tone or noise input.

13. The voice processing system of any of claims 7 to 12, further comprising means for calculating one or more acoustic parameters of the audio input to assist determining whether to continue or to terminate playing out of said prompt.

14. A computer readable medium containing computer program instructions for a voice processing system for providing speech recognition with barge-in, said computer program instructions comprising instructions for:

playing out a prompt to a user;

receiving audio input from the user whilst said prompt is still being played out;

performing speech recognition on said audio input to determine a corresponding text;

performing lexical analysis on said text to determine whether or not the text satisfies one or more conditions;

and responsive to said text satisfying said one or more conditions, terminating the playing out of the prompt; otherwise, continuing the playing out of said prompt.

VOICE PROCESSING SYSTEM

ABSTRACT

A voice processing system includes a speech recognition facility with barge-in. The system plays out a prompt to a caller, who starts to provide their spoken response whilst the prompt is still being played out. The system performs speech recognition on this response to determine a corresponding text, which is then subjected to lexical analysis. This tests whether the text satisfies one or more conditions, for example, including one or more words from a predefined set of task words. If this is found to be the case, the playing out of the prompt is terminated (ie barge-in is effected); otherwise, the playing out of the prompt is continued, essentially as if the caller had not interrupted.

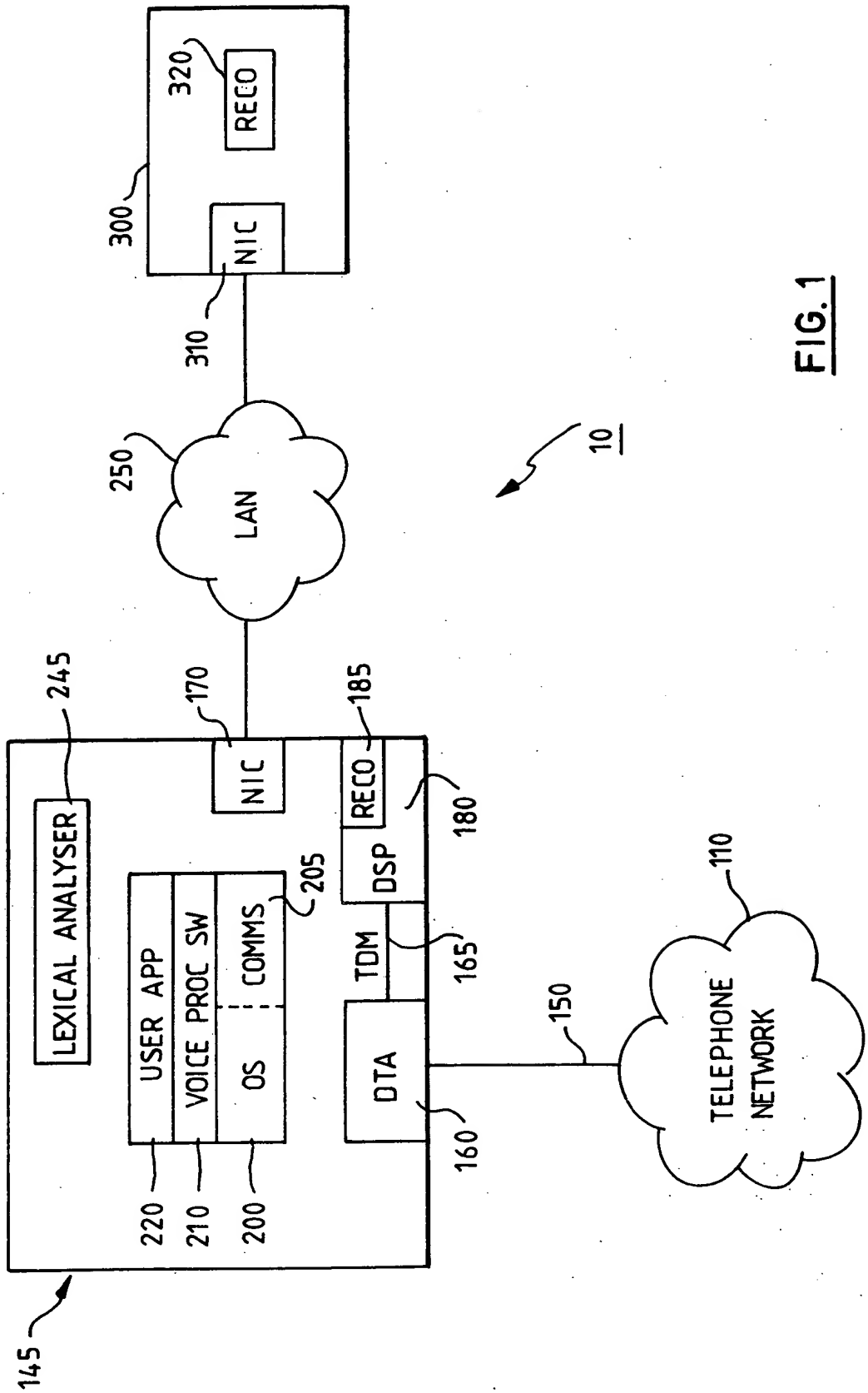
300
320
RECO
NIC
310
250
LAN
10
245
170
185
LEXICAL ANALYSER
NIC
RECO
180
205
USER APP
VOICE PROC SW
COMMS
OS
DSP
TDM
165
150
110
TELEPHONE NETWORK
DTA
160
220
210
200
145

FIG. 1

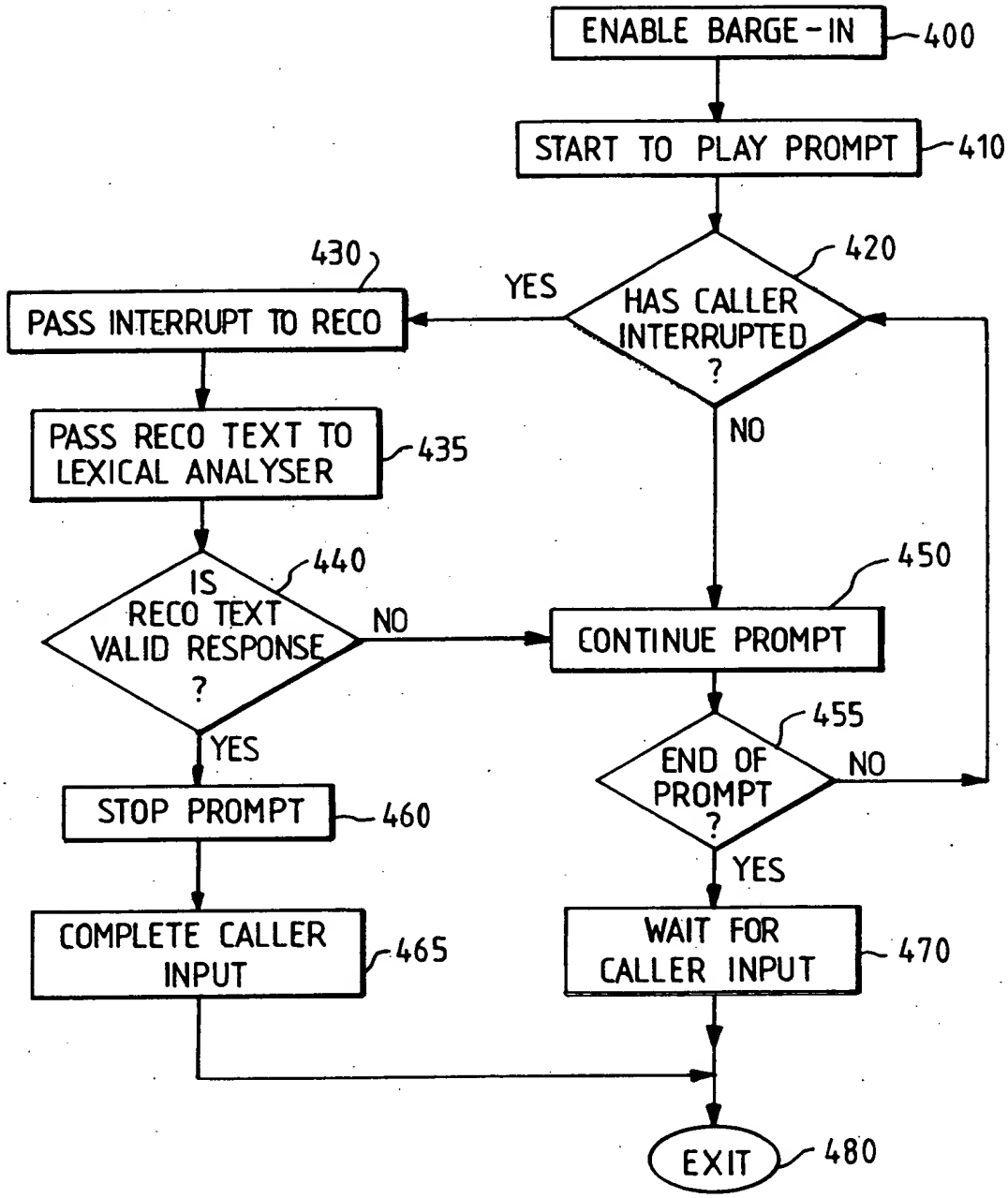
ENABLE BARGE-IN
400
START TO PLAY PROMPT
410
HAS CALLER INTERRUPTED ?
420
YES
NO
430
PASS INTERRUPT TO RECO
PASS RECO TEXT TO LEXICAL ANALYSER
435
IS RECO TEXT VALID RESPONSE ?
440
NO
YES
CONTINUE PROMPT
450
END OF PROMPT ?
455
NO
YES
STOP PROMPT
460
COMPLETE CALLER INPUT
465
WAIT FOR CALLER INPUT
470
EXIT
480

FIG. 2